УДК 681.3.06:504.009

ВЫЧИСЛИТЕЛЬНЫЕ ТЕХНОЛОГИИ В ЗАДАЧАХ ОБРАБОТКИ ДЕНДРОЭКОЛОГИЧЕСКИХ ДАННЫХ

И.А. Ботыгин, Ю.В. Волков, В.Н. Попов, В.А. Тартаковский*

Томский политехнический университет
*Институт мониторинга климатических и экологических систем СО РАН. г. Томск
E-mail: botygin@ad.cctpu.edu.ru vnp@mail2000.ru tv@iom.tomsknet.ru

Представлена вычислительная технология, обеспечивающая доступ к информационным материалам в области дендроэкологии и программным средствам для анализа изображений годичных колец деревьев с целью выявления изменений параметров окружающей среды, отраженных в приросте дерева. В основу моделирования годичных колец деревьев положены фундаментальные свойства радиального роста дерева - монотонность роста во времени и пространстве и ограниченность его скорости, а изменения плотности древесины рассматриваются как пространственно-временной колебательный процесс. Технология базируется на динамическом управлении распределёнными вычислительными ресурсами.

1. Введение

При реализации достаточно сложных задач очень быстро проявляется недостаток имеющихся вычислительных мощностей, заключающийся, как правило, в увеличении времени отклика системы. В частности, особенность задач дендроэкологического анализа заключается в необходимости математической обработки очень большого объема данных (иногда требуется обработка временных рядов наблюдений за 1000 лет и более), широким спектром решаемых задач, коллективной (одновременной) работой многих сотрудников, практически, на всех этапах дендроэкологических исследований, а также в необходимости хранения и систематизации больших объемов неоднородной структурированной информации (собственно хронологические ряды наблюдений, результаты обработки, сопутствующие метеорологические, геологические, геофизические, аэрокосмические и т.п. ряды наблюдений).

Отмеченные выше особенности задач дендроэкологии не позволяют создавать профессиональные центры обработки и хранения дендроэкологических данных в отдельно взятом научном учреждении. Слишком большие вычислительные возможности и мощные базы данных требуются для этого и слишком велики материальные затраты. Естественным решением в сложившейся ситуации является распределение информационного, программного и технического обеспечения (ресурсов) решения задач дендроэкологии в нескольких специализированных центрах обработки и виртуализация доступа к этим информационно-обрабатывающим ресурсам исследователей. Такая концепция однозначно подразумевает в качестве одного из системных решений использование в таких профессиональных центрах технологии распределенной обработки (*grid*-технологии), обеспечивающей динамическое изменение основных компонентов инфраструктуры системы обработки (от структур хранимых данных – до схем и алгоритмов решаемых задач) и повышение вычислительной мощности.

Наиболее значимым при решении задач дендроэкологии становится восстановление ретроспективных сведений об изменчивости условий окружающей среды, при изучении которых используются результаты, полученные путём инструментальных измерений. Такая информация является наиболее ценной и широко используемой, но часто обладает существенными недостатками, которые ограничивают её использование, в частности, кратковременность и нерегулярность проводимых наблюдений, их территориальная ограниченность и т.д. В связи с этим перспективным подходом к реконструкции хода природных процессов являются индикационные методы, основанные на наличии связей между составом и структурой природных тел и параметрами внешней среды. Индикаторами в таких методах обычно являются объекты живой природы, характеризующиеся неоднородными анатомо-морфологическими характеристиками, возникшими в разные годы, сезоны года и даже в течение суток, и сохраняющими такое строение в течение длительного времени. Наиболее известным и широко распространенным примером регистрирующих структур служат годичные слои (годичные кольца) древесины в стволах, ветвях и корнях деревьев и кустарников [1–3]. Отметим, что годичные кольца дают интегральную информацию о качестве жизни и в этом их преимущество перед физическими методами измерения параметров окружающей среды, которые могут с высокой точностью определить значения конкретных параметров (концентрацию веществ, интенсивность излучения, скорость движения, температуру и прочее). Однако живому организму важно их оптимальное сочетание, что и фиксируют годичные слои древесины [3, 4].

Таким образом, исследование особенностей роста дерева (от начала до завершения) с детальным изучением ростовой активности внутри сезона с использованием развитого математического описания изменения профиля плотности древесины, поддержанное численными алгоритмами, реализованными в высокопроизводительных вычислительных системах с новыми эффективными распределёнными методами обработки дендроэкологических данных, даст возможность получить новые результаты и качественные оценки экологических параметров окружающей среды.

2. Сравнительный обзор и анализ аппаратно-программных средств обработки и анализа годичных колец деревьев

Анализ научных исследований в области дендроэкологии, дендрохронологии и дендроклиматологии показывает, что в настоящее время для моделирования и анализа дендроэкологических данных используется достаточно широкий спектр аппаратно-программных средств [5, 6]. Использование программных средств зависит от задач исследования. Это может быть статистическая обработка первичных данных замеров характеристик прироста и получение надёжных обобщённых хронологий, сопоставление характеристик прироста с факторами внешней среды, моделирование процессов роста, изучение пространственного распределения характеристик прироста и визуализация результатов анализа. Для решения этих задач одни исследователи предпочитают использовать широко распространённые статистические пакеты (SAS, MatLab, SyStat, STATISTICA и др.), электронные таблицы (QUATTRO, LOTUS, Excel и др.), универсальные системы математической обработки результатов, предназначенные для численного и символьного решения математических задач различной сложности (MathCAD, Mathematica и др.) и интегрированные программные решения такие как DPL, PRECON 5.17C, TREERING 3.0, LignoVision 1.32, TSAP-Win Professional 0.30, DendroClim 2002, WinDENDRO, DendroLab 470, PAST 32, OSM 3.10, MeasureJ2X, или решения на базе географических информационных систем (ESRI ARC/INFO, ArcView, Mapinfo, AutoCAD Map и др.) с форматами различных баз данных (Microsoft Access, Oracle, dBASE, FoxPro и др.) для связи обрабатываемых данных с географическим местоположением их сбора и отображением на электронных картах [3].

Рассмотренные программные комплексы и системы достаточно универсальны, ориентированы на широкий класс исследований и выполнение часто используемых операций обработки и анализа дендроэкологических данных (например, измерение ширины годичных колец и др.). Однако, они плохо учитывают специфику требований конкретных исследователей (например, определение азимута и среднеквадратичной ширины области максимального прироста и др.).

3. Монотонность и непрерывность роста как основа построения математической модели годичных колец

Математическая модель радиального сечения дерева строится, исходя из формы и структуры концентрических годичных слоёв, которые формируются на поперечном спиле дерева в процессе его роста. Для построения модели вводится полярная система координат, начало которой совпадает с биологическим центром диска дерева. Годичные кольца рассматриваются как функция двух переменных: ρ (радиуса) и θ (азимута). На рис. 1 показано поперечное сечение ствола дерева (пихты сибирской, *Abies sibirica Ledeb*), полученное в результате сканирования с необходимой дискретностью, и профиль изменения интенсивности окраски вдоль радиуса сканирования (функция $R(\rho)$, отмеченной пунктиром). На графике значения функции $R(\rho)$ представлены в относительных единицах, а аргументом является расстояние вдоль радиуса сканирования. Заметим, что фрагмент значений функции $R(\rho)$ представлен по наиболее наглядному участку радиуса сканирования.

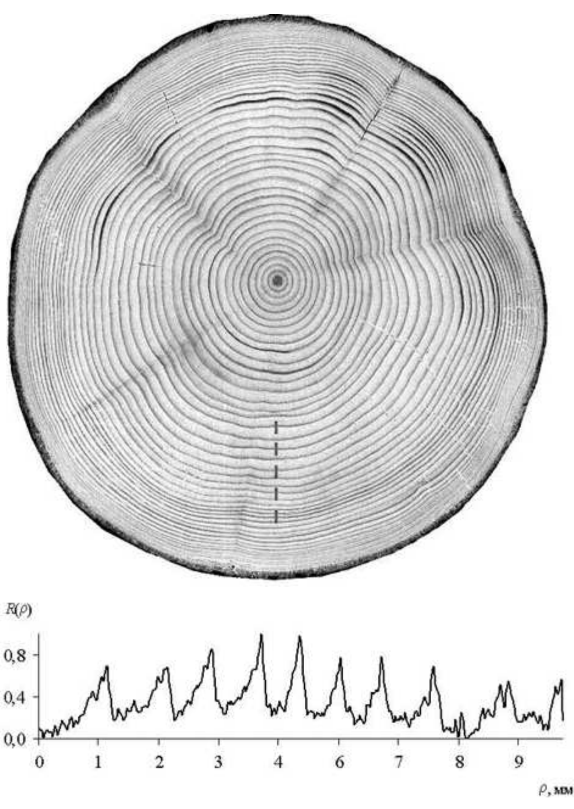


Рис. 1. *Радиальное сечение диска дерева*

Принимаем функцию $R(\rho)$ близкой функции плотности древесины и будем называть её функцией плотности. Функция $R(\rho)$ представляет собой колебательный процесс, период которого (период радиального роста дерева) непостоянен. В конце каждого периода, на границе между поздней и ранней древесиной, происходят резкие изменения функции плотности, которыми и фиксируется время (τ) при соответствующих значениях ρ, что фактически определяет на радиусе временные реперы. Вследствие того, что радиальный рост не может убывать и меняться скачком во времени в пределах вегетационного периода, считаем, что функция роста $\rho(\tau)$ является монотонной и имеет ограниченную производную. По этой причине обратная ей функция $\varphi(\rho)$ также монотонна и, более того, она является строго возрастающей функцией.

Из экспериментальных данных следует, что сечение $R(\rho)$ содержит шумы, наложенные на основное колебание. Локальный импульсный шум $h(\rho)$ определён в области ранней древесины. Этот шум вызван главным образом недостаточной чистотой

обработки поверхности поперечного спила и естественными флуктуациями размера клеток древесины. В виду того, что формирование клеток происходит под действием большого числа различных факторов, данный компонент имеет случайный характер. Локальный импульсный шум занимает внутреннюю область годичного кольца, имеет малую амплитуду, по отношению к амплитуде полезного сигнала, и расположен в области высоких частот. Наряду с высокочастотным шумом, имеется глобальный низкочастотный шум $l(\rho)$, связанный с наличием возрастных зон на поперечном спиле ствола дерева: сердцевины и заболони. Принимаем, что этот компонент шума накладывается на полезный сигнал в виде колебания, период которого значительно больше периода, соответствующего наибольшему годичному кольцу.

Исходя из вышесказанного математическую модель радиального сечения годичных колец можно представить в виде суммы шумов и действительного колебания $U(\rho)$, модулированного по амплитуде и фазе:

$$R(\rho)=C+U(\rho)+h(\rho)+l(\rho)>0, \quad (*)$$

$$U(\rho)=a(\rho)\cos\varphi(\rho),$$

$$\varphi(\rho)=\frac{2\pi}{T}\tau(\rho),$$

где $a(\rho)$ – амплитуда колебания, $\varphi(\rho)$ – фаза колебания, $\tau(\rho)$ – время роста, T – период вегетации, C – постоянная.

Функция $R(\rho)$ формируется биологическим объектом под влиянием внешней среды и имеет вид интерференционной картины. Поэтому естественно предположить, что функция плотности древесины является результатом суммарного взаимодействия “биологической” и “экологической” составляющих. Отметим, что подобная формализация радиального сечения диска дерева позволяет эффективно применить её для анализа как внешних, так и внутренних особенностей роста дерева в вегетационный период. Основой для проведения прикладного анализа роста являются свойства компонентов, входящих в построенную математическую модель. Главной и наиболее информативной составляющей модели является фаза, так как она связана с функцией роста обратно пропорциональной зависимостью.

Для восстановления роста из выражения (*) находится фаза, а затем определяется обратная ей функция. После последовательных преобразований функция роста будет иметь вид:

$$\rho(\tau)=\tau^{-1}\left(\frac{T\varphi}{2\pi}\right).$$

Функция роста $\rho(\tau)$ даёт время появления каждой клетки дерева на исследуемом радиусе в процессе вегетации, т.е. для всех значений ρ в интервале между реперами. Полученное аналитическое выражение уже пригодно для создания алгоритмов автоматического вычисления годового прироста деревьев за вегетационный период.

Дальнейшие уточнения модели радиального сечения диска дерева связаны с различными методами определения амплитуды и фазы. В частности, широкое распространение для вычисления амплитуды и фазы получил аналитический сигнал, широко используемый в теории колебаний и волн. Аналитический сигнал является комплексным, его действительная часть есть исходная функция, а мнимая часть образуется применением преобразования Гильберта к действительной части. Аналитический сигнал определяет способ измерения амплитуды, фазы и угловой частоты сигнала. Если спектр сигнала сосредоточен в относительно узком интервале частот (квазимонохроматический сигнал), то изменения амплитуды и фазы за время, соответствующее периоду основной частоты, при определенных условиях оказываются минимальными [7].

4. Структура и программная реализация многомашинного комплекса обработки дендроэкологических данных

Авторами были проведены экспериментальные исследования требуемой вычислительной мощности на такой задаче в области дендроэкологии как обработка изображений годичных колец деревьев (сканирование, обработка фильтрами). В зависимости от размера обрабатываемого изображения, требовалось до нескольких дес. мин процессорного времени (Celeron 1,7 МГц, DDR RAM 256 Мб). Нагрузка на процессор возрастала, практически, пропорционально числу подключившихся пользователей. Одним из решений проблемы недостаточности вычислительных ресурсов является кластеризация нескольких компьютеров в многомашинный комплекс обработки (рис. 2), так как реально увеличивать вычислительную мощность одного компьютера до бесконечности проблематично.

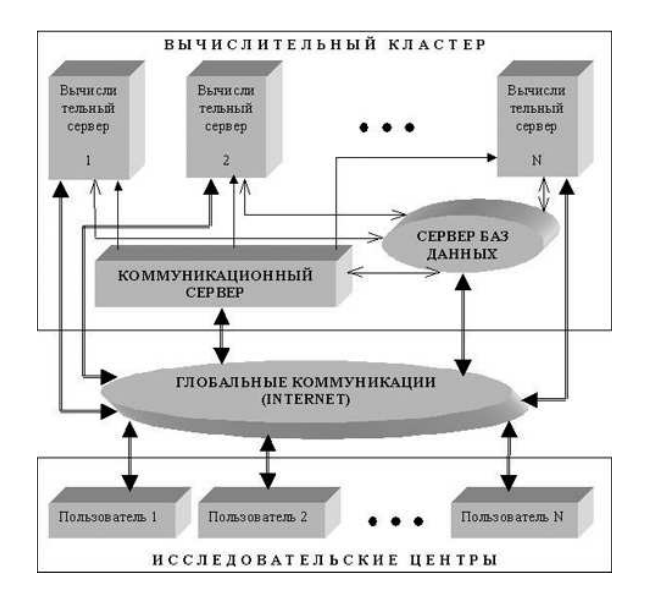


Рис. 2. *Архитектура комплекса обработки дендроэкологических данных*

Планирование и диспетчирование процессами обработки дендроэкологических данных возложено на специальный коммуникационный сервер. Взаимодействие всех пользователей с комплексом обработки осуществляется только через коммуникационный сервер. Непосредственно на этом сервере математические вычисления не выполняются. Основной задачей коммуникационного сервера является обеспечение оптимальной загрузки имеющихся в его распоряжении вычислительных серверов и обеспечение режима работы в реальном времени (*online*) пользователей. Решение задач математической обработки данных (например, таких как вычисление радиального роста дерева и плотности древесины в период вегетации, проведение спорово-пыльцевого анализа последовательных слоёв торфа или построение длительных древесно-кольцевых хронологий и т.д.) возложено на вычислительные серверы кластера.

Для информационной поддержки функционирования многомашинного комплекса обработки дендроэкологических данных использована СУБД Oracle. С сервером базы данных Oracle взаимодействуют как вычислительные серверы, так и коммуникационный сервер. Вычислительные серверы кластера, реализующие основную математическую обработку, и коммуникационный сервер выполнены на базе Apache HTTP Server, а клиентская часть, реализующая интерфейс удалённого взаимодействия с пользователем, использует MS Internet Explorer.

Рассмотренная выше схема кластеризации ресурсов была реализована для решения задачи анализа изображений годичных колец деревьев. Вычислительный кластер обеспечивал решение таких задач, как сбор, систематизацию и хранение научной информации в области дендроэкологических исследований, формирование и ведение базы экспериментальных дендроэкологических данных, вычисление азимута и среднеквадратичной ширины области максимального прироста годичных колец деревьев, вычисление значений индексов прироста ширины годичных колец деревьев, оценку экологического состояния окружающей среды. Для указанных задач были разработаны и реализованы авторские вычислительные алгоритмы. В качестве критерия, определяющего качество алгоритмов и их программной реализации, было выбрано время выполнения. Оптимизация осуществлялась по минимуму времени выполнения [8].

Ниже приведена иллюстрация функционирования комплекса при решении некоторых из перечисленных выше задач. В задаче вычисления азимута и среднеквадратичной ширины области максимального прироста годичных колец деревьев исследование включало: определение центра колец, сканирование изображения по всем азимутам, преобразование изображения из полярной системы координат в декартовую, обработку изображения медианным (рис. 3) и полиномиальным (рис. 4) фильтрами, преобразование изображения из градаций серого в чёрно-белое, расчёт ширины годичных колец и интегрального прироста по всем азимутам, нормирование ширины прироста каждого кольца.

Процедура медианной фильтрации изображения использовалась для устранения импульсного шума. В результате медианной фильтрации происходит ограничение сигнала сверху и снизу отрезками, параллельными оси абсцисс. Это проиллюстрировано на рис. 3, где изображена функция плотности древесины до и после применения медианной фильтрации с пятиэлементным окном. Так как исходная функция плотности древесины, наряду с импульсным шумом, содержит и низкочастотный шум, то его устранение осуществляется фильтрацией в полиномиальном базисе на основе метода наименьших квадратов. Из рис. 4 видно, что функция плотности древесины после применения полиномиальной фильтрации (полином пятой степени) не ухудшила своего вида, а низкочастотный шум исчез. В результате фильтрации получается сигнал, из которого выделена низкочастотная составляющая и импульсный шум при сохраненной информации о переходных зонах и незначительном изменении амплитуды.

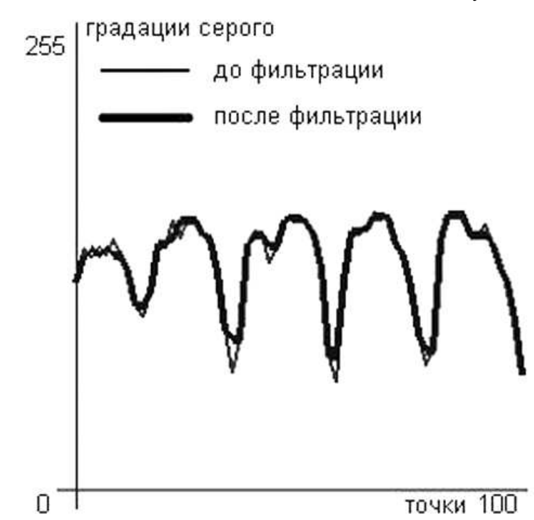


Рис. 3. *Медианная фильтрация сигнала*

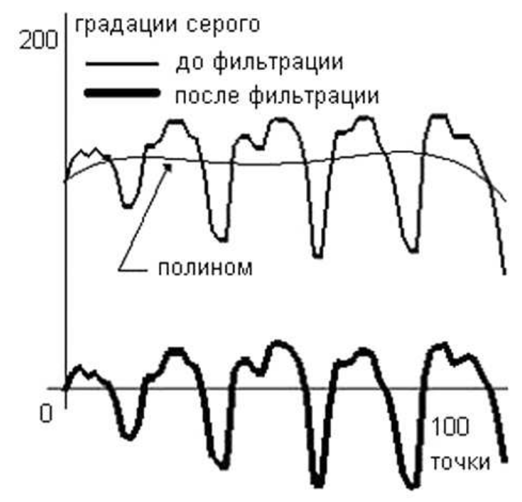


Рис. 4. *Полиномиальная фильтрация сигнала*

Также проводились вычисления значений индексов прироста по ширине годичных колец деревьев – сканирование изображения по выбранному сечению, обработка полиномиальным фильтром, преобразование из градаций серого в чёрно-белое, подсчета средних значений по индексам прироста годичных колец, рис. 5.

На графике по оси *x* отложены года, по оси *y* – прирост и значение индекса. Исходная ширина годичных колец деревьев отмечена квадратами в узловых точках, сглаживающая кривая – вертикальными прямыми, индексы ширины – окружностями.

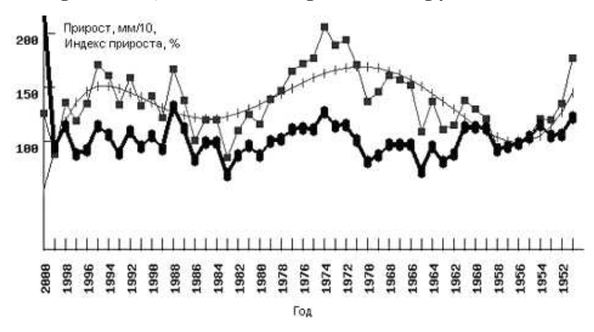


Рис. 5. *Вычисление значений индексов прироста годичных колец деревьев*

В данном случае индекс прироста – трансформированное значение ширины годичных колец, полученное путём деления её на соответствующее ей значение на сглаженной кривой. Вычисление значений индексов прироста осуществляется с целью устранения возрастного тренда из исходных измерений путём перевода последовательности значений прироста в безразмерные значения индексов ширины годичных колец. В изменчивости индексов влияние внешних условий максимально и процедуры дальнейшего анализа позволяют выделить ведущие факторы среды и оценить их относительный вклад.

В целом, полученные оценки позволяют производить реконструкцию многих важных климатических характеристик, основными из которых являются температура воздуха в различные сезоны и за год, количество осадков в различные сезоны и за год, аномалии атмосферного давления, повторяемость и интенсивность засух, повторяемость заморозков в течение вегетационного периода и сильных морозов в зимнее время.

Заключение

Представлена архитектура многомашинного комплекса, обеспечивающая динамическое изменение инфраструктуры системы обработки и доступ к программным средствам математического анализа годичных колец деревьев с целью мониторинга окружающей среды. Математическая обработка основывается на представлении плотности древесины как пространственно-временном колебательном процессе и включает вычисление азимута и среднеквадратичной ширины области максимального прироста, индексов прироста ширины годичных колец деревьев.

Работа выполнена при поддержке Российского фонда фундаментальных исследований, проект «Обь» 05-07-98008.

СПИСОК ЛИТЕРАТУРЫ

1. Биоиндикация загрязнений наземных экосистем / Под ред. Р. Шуберта – М.: Мир, 1998. – 350 с.
2. Розенберг Г. С. Биоиндикация: теория, методы приложения. – М.: Наука, 1983. – 163 с.
3. Шиятов С.Г., Ваганов Е.А., Кирдянов А.В., Круглов В.Б., Мазепа В.С., Наурзбаев М.М., Хантемиров Р.М. Методы дендрохронологии. Ч. I. Основы дендрохронологии. Сбор и получение древесно-кольцевой информации. – Красноярск, 2000. – 80 с.
4. Ваганов Е.А., Шашкин А.В. Рост и структура годичных колец хвойных. – Новосибирск: Наука, 2000. – 232 с.
5. Ultimate Tree-Ring Web Pages. Software used in Dendrochronology. Обращение к документу 23.06.2004. http://web.utk.edu/~grissino/software.htm
6. Экологическая биофизика / Под ред. И.И. Гительзона, Н.С. Печуркина. Т. 2. Биофизика наземных и водных экосистем. – М.: Логос, 2002. – 360 с.
7. Тартаковский В.А., Исаев Ю.Н., Несветайло В.Д., Волков Ю.В., Попов В.Н. Математическая модель радиального сечения годичных колец деревьев // Автометрия. – 2003. – Т. 38. – № 5. – С. 118–127.
8. Botygin I.A., Popov V.N., Volkov Ju.V., Tartakovsky V.A. Information technology of tree-ring processing // Proc. of the 8th Korea-Russia Intern. Symp. on Science and Technology. – KORUS-2004. – Tomsk, 2004. – V. 1. – P. 15–18.